

日期: 2021/01/12(二) 上機考: Open Book
授課教師: 吳漢銘 (臺北大學統計學系副教授)

請仔細閱讀每一個注意事項 (禁止討論)

1. 考試期間

- (a) 請按照平時上課之座位入座。若同學右邊有人，則該同學應使用隔板 (或拿包包擋著)。
- (b) 可參考課本、上課講義 (包含電子檔) 及其它資料，但不能與別人討論。
- (c) 可使用計算機、自己的筆記型電腦及平板電腦，不可使用手機。
- (d) 全程可上網查詢，但不能用通訊軟體 FB/LINE 等討論。
- (e) 程式設計題，若程式碼直接複製 (或照抄) 講義上的以不給分為原則。
- (f) 有問題者，請舉手發問。勿與同學交談。
- (g) 不按照規定作答者，酌量扣分。
- (h) 請參照下列文件第 2 ~ 4 頁寫作規定，不按照規定作答者，會扣分。
<http://www.hmwu.idv.tw/web/teaching/doc/R-how-homework.pdf>
- (i) 不可使用它人之隨身碟。「作弊」或「疑似作弊」，往後各項考試不予評分。
- (j) 程式請隨時存檔，避免突然意外發生，程式檔不見。

2. 下載題目卷，上傳答題檔案:

- (a) 於課程網站下載題目卷。
- (b) 上傳答題檔案: 於課程網站上登入 [作業考試上傳區]，帳號: r1091。密碼: xxxxx。
- (c) 上傳答題檔案時，請注意「正確目錄」。
- (d) 答題是從「Console」複製程式執行過程及結果，然後貼到答題案。
- (e) 貼圖時，圖上之字型、符號需可辨認。圖形大小適中，不可變形。
- (f) 若傳錯，請最終要上傳一份正確的答題檔案。
- (g) 請上傳「學號-姓名-R-FinalExam.docx」。(學號及姓名，改成自己)
- (h) 若上傳檔案格式錯誤，內容亂碼，空檔等等問題。請自行負責。
- (i) 若要重覆上傳 (第 2 次以上)，請在檔名最後加「-2」、「-3」，例如: 「學號-姓名-R-FinalExam-2.docx」等等。
- (j) 上傳兩次 (含) 以上、格式不合、於規定時間之後才上傳等等酌量扣分。

3. 完成考試: 上傳完畢，確認檔案大小無誤，關閉電腦後，即可離席。

我已經仔細閱讀上述各注意事項，若有違背，會自行負責。

日期: 2021/01/12(二) 上機考: Open Book
授課教師: 吳漢銘 (臺北大學統計學系副教授)

1. 一袋中有 5 顆紅球及 3 顆白球，小明由袋中隨機抽球，每次取一球，共取 4 次，令 A 為抽出 2 次白球的事件，計算此事件分別在放回 (replacement)、不放回 (without replacement) 兩種情況下之機率 $P(A_r), P(A_w)$ 。

Solution:

$$\text{放回: } P(A_r) = C_2^4 \left(\frac{3}{8}\right)^2 \left(\frac{5}{8}\right)^2$$

$$\text{不放回: } P(A_w) = \frac{C_2^5 C_2^3}{C_4^8}$$

- (a) 請直接計算上述之機率 (分別命名為 `Prob.Ar`, `Prob.Aw`) 並印出。(註: C 為組合數, R 指令為 `choose`)
- (b) 小明今天想要以寫 R 程式的方式來模擬此隨機實驗，計算抽球的機率。若設定 `set.seed(123456)`，列出「一袋中有 5 顆紅球及 3 顆白球，小明由袋中隨機抽球，分別在放回 (replacement)、不放回 (without replacement) 兩種情況下，每次取一球，共取 4 次」實驗一次的結果，並計數印出白球出現之個數。(不需寫成 R 函式)
- (c) 同上小題，寫一 R 函式 (命名 `Draw_Result`)，輸入為「是否印出每次實驗結果 (`is.print`)」，預設值為 `TRUE`，輸出為在兩種情況下，白球各自出現之個數。印出重覆此實驗 10 次的結果。
- (d) 同上小題，重覆上述實驗 1000 次，不印出每次結果，分別在放回、不放回兩種情況下，計算抽中 2 顆白球機率。

2. 大一線性代數上學期期末考有一證明題:

Let A be an $m \times n$ matrix. Prove that $\text{rank}(A) = \text{rank}(A^T A)$ 。

今天小明很無聊，想用 R 程式來驗證一下，他的做法如下:

- 先寫一 R 函式，命名為 `Check_Rank`，輸入為 `m`, `n`，輸出為一資料框 (`data.frame`)，欄位為英文命名，依序為: `no(次數)`, `m`, `n`, `rankA`, `rankATA`。
- 在 `Check_Rank` 函式內，隨機產生任一 $m \times n$ 矩陣，隨機數來自標準常態分佈: `x <- rnorm(n*m)`。
- 查詢在 R 如何計算一矩陣的 Rank。
- 以 `seq` 指令產生 `m = 3, 5, 7, 9, 3, 5, 7, 9, 3, 5, 7, 9` 及 `n = 4, 4, 4, 4, 8, 8, 8, 8, 12, 12, 12, 12`，共 12 次組合當測試。

現在請你幫他實現這個做法。

3. 某商業公司舉行抽獎活動，中獎名單如下，因考量個資法，公告名單不能將全名公開，請你幫此名單，每一中獎者的姓名及會員卡號，部份字元打上「{tt *}」，例如:「沈 * 予 7113***607」，印出修改後可公告之名單。

	會員姓名	會員卡號
1	沈俞予	7113235607
2	簡惠榕	8010785376
3	徐一良	9010344896
4	賴淨茹	3010789872
5	林金玲	5011213845
6	吳彩鳳	2592903839
7	江德翰	3714483694
8	葉建鴻	4012123657
9	阮通全	3053398421
10	黃玉鈴	3317005422

4. 資料集檔案 `WaterQuality_Hsinchu.xlsx` 為「新竹市河川水質」近一年「監測資料」之紀錄，主要欄位為: 河川名稱、測站名稱、採樣日期、採樣時間、溶氧量 (DO)、生化需氧量 (BOD5)、懸浮固體 (SS)、pH 值 (pH)、導電度 (EC)、水溫 (WT)、大腸桿菌群 (Coli_G)、氨氮 (NH3_N)、鉛 (Pb)、六價鉻 (Cr6)、鎘 (Cd)、銅 (Cu)、鋅 (Zn)、鎳 (Ni) 等等。請參見網址: <https://data.gov.tw/dataset/67604>。

- (a) 讀取此資料集檔案，印出此資料集之紀錄筆數及欄位個數，並印出所有欄位名稱。
- (b) 產生一具有三個類別的因子 (順序) 變數: 水溫組 (命名為 `WT.group`)。規則如下: 若水溫 (WT) (°C) 高於平均的 2 倍標準差，則 `WT.group` 為高溫組，若水溫 (WT)(°C) 低於平均的 2 倍標準差，則 `WT.group` 為低溫組，其餘為均溫組。印出三個類別的個數 (提示: `ifelse`, `table`)
- (c) 畫出下列四個變數之索引圖: 溶氧量、pH 值、大腸桿菌群、氨氮。每一張圖之點符號 (symbol) 顏色為水溫組之絡類別。(註: 一頁 4 張圖，每張圖的標題為變數名稱及其單位)
- (d) 畫出下列四個變數之直方圖: 溶氧量、pH 值、大腸桿菌群、氨氮。(註: 一頁 4 張圖，每張圖的標題為變數名稱及其單位)